Measurement Invariance of R-UCLA Loneliness Scale Across Gender Using Differential Item Functioning (DIF)

Mahnaz Shojaee[1,*], Mehrdad Shahidi[2,3], Ying Cui[1], Amin Mousavi[4]

[1]Centre for Research in Applied Measurement and Evaluation, University of Alberta, Canada
[2]Department of Education, Mount Saint Vincent University, Canada
[3]Department of Psychology, Tehran Central Branch, Islamic Azad University, Tehran, Iran
[4]Department of Educational Psychology and Special Education, College of Education, University of Saskatchewan, Canada

Abstract Background: Using R-UCLA-Loneliness Scale, previous studies across Canada and other Western and Asian countries indicated that loneliness could predict some of the mental health disorders such as clinical depression, anxiety, as well as physical problems. Thus, the self-report R-UCLA-Loneliness Scale has been used increasingly to screen and monitor the sense of loneliness in people with and without mental health disorders. A brief review of previous research on the R-UCLA-Loneliness Scale revealed the lack of evaluation of the scale's fairness/equivalence in measuring loneliness among adolescents and young adults. Method: Four hundred university students aged between 18 and 26 years old participated in this study. Their responses were analyzed using item response theory (IRT) and ordinal logistic regression (OLR). To determine the type of Differential Item Functioning (DIF), uniform or non-uniform, we examined "item true score function plots". Results: The findings revealed that three items display gender-related DIF. All three items showed non-uniform DIF. The assessment of effect sizes indicated negligible differences for practical utilization. Limitations: A caution should be considered in interpreting the results of measurement invariance across Gender. The samples analyzed in this study were individuals without comorbid symptoms or other mental disorders. Thus, we recommend future researchers focus on the clinical population. Conclusion: This study was the first and major step towards presenting theoretical and practical information regarding the assessment of invariance of the R-UCLA-Loneliness Scale across gender. Future studies may look at different aspects and sources of invariance such as marital status and educational levels for different groups to strengthen conclusions concerning the R-UCLA Loneliness Scale.

Keywords Loneliness, Differential Item Functioning (DIF), Item Response Theory (IRT), Measurement Invariance, R-UCLA Loneliness Scale, Gender

1. Introduction

Loneliness is one of the most prevalent psychological problems across Canada (Kaiser Family Foundation, 2020; Keefe, Andrew, Fancey, & Hall, 2006) and around the world (Groarke, Berry, Graham-Wisener, McKenna-Plumley, McGlinchey, & Armour, 2020; Igbokwe, Ejeh, Agbaje, Umoke, Iweama, & Ozoemena, 2020; McGinty, Presskreischer, Han, & Barry, 2020). As a unitary and global phenomenon, loneliness is characterized by the lack of emotional bonds with others, lack of caring people, a sense of being socially isolated, a sense of being silly, empty, upset, not right, disappointed, confused, sad, and bored (Layden, Cacioppo, & Cacioppo, 2018; Masi, Chen, Hawkley, & Cacioppo, 2011; McWhirter, 1990; Russell, McRae, & Gomez, 2012). These symptoms can directly affect

individuals' psycho-social functions (Cacioppo & Hawkley, 2009; Layden, Cacioppo, & Cacioppo, 2018).

Having close association with social isolation, loneliness in a unidimensional model is considered as a subjective perception of social isolation (Ercole & Parr, 2020). However, social isolation indicates an objective measure of individual's social bonds or social relationships (Ercole & Parr, 2020; Russell et al., 2012; Smith, & Victor, 2019). Thus, in some situations in which individuals' interactions are limited seriously such as compulsory social distance during the coronavirus 2019 (COVID-19) pandemic, the loneliness increasingly influences individual's functions in a negative way (Cauberghe, Wesenbeeck, De Jans, Hudders, & Ponnet, 2020; Holmes et al., 2020). In such situations, individuals may show a sense of loneliness along with the symptoms of affective disorders such as the feelings of sadness, loss of interest, emptiness, depressed, hopeless, or confused (Cauberghe et al., 2020; Groarke et al., 2020).

Several findings indicated that loneliness could predict other mental health disorders such as clinical depression (Buchholz & Catton, 1999; Jaremka, Fagundes, Glaser,

Bennett, Malarkey, & Kiecolt-Glaser, 2013; Wakefield, Bowe, Kellezi, Butcher, & Groeger, 2020; Wiseman, Mayseless, & Sharabany, 2006) and anxiety (Igbokwe et al., 2020). As well, it can affect physical functions negatively and cause problems such as asthma, heart disease, cancer, diabetes, and shortness of breath (Ercole et al., 2020; Groarke et al, 2020).

Accordingly, clinical practitioners, psychiatrists, and psychotherapists endeavor to screen and recognize loneliness as early as possible by using reliable and valid tools for further therapeutic intervention. A recent review of most frequently used loneliness measures by Suri and Garg (2020) revealed that among different unidimensional scales, R-UCLA Loneliness Scale (3rd version), originally constructed and developed by Russell, Peplau, and Ferguson (1978; Russell, Peplau, & Cutrona, 1980), is one of the most appropriate scales with several vital characteristics for measuring loneliness such as: 1) easy administration, 2) sufficient number of questions, 3) self-report rather than clinician administered, 4) globally used with acceptable internal reliability and validity, and 5) ability to measure change overtime (Groarke et al., 2020; Russell, Peplau, & Cutrona, 1980; Shahidi, 2013; Suri et al., 2020). Since 1978, the time UCLA Loneliness Scale was introduced, numerous studies have been conducted around the world to either examine the psychometric properties of the scale or to study loneliness in relation to other mental health conditions in diverse populations (Groarke, et al., 2020; Igbokwe et al., 2020; Russell et al., 1978; Russell et al., 1980; Russell et al., 1996; Russell et al., 2012; Shahidi, 2013; Suri, et al., 2020).

A brief review of previous studies on R-UCLA-Loneliness Scale (3rd Version) revealed the lack of evaluation of R-UCLA-Loneliness Scale's (3rd Version) fairness/equivalence in measuring the sense of loneliness among identified groups specially males and females. Also, there were differences between females and males in terms of their scores on the loneliness scale (Shahidi, 2013). However, there was not obvious whether such differences (e.g., females were lonelier than males) were the results of item biases in the scale or of the nature of gender diversity. Accordingly, since assessing measurement equivalence is an important part of the process of validating scales to test whether the probability of responding to a specific item exhibits different statistical properties for different identifiable groups after matching the groups on the construct being measured (De Ayala, Kim, Stapleton, & Dayton, 2002; Sharafi, Mousavi, Ayatollahi, & Jafari 2017), the aim of this study was to assess the measurement equivalence of loneliness by means of R-UCLA version three across gender to ensure about its fairness/equivalence.

2. Method

2.1. Participants

Employing a systematic multistage random sampling method in the Islamic Azad University-Tehran Central Branch in Iran, 400 university students were selected. This sample had a mean age of 21.66 (SD = 2.20) with a minimum of 18 and a maximum of 26 years old. Of these participants, 361 (90%) individuals were single, 36 (9%) were married and 3 (0.8%) were divorced. The gender composition in this sample was 200 (50%) females and 200 (50%) males. About half of the participants were studying in Psychology and the other half in other disciplines.

2.2. Measures

2.2.1. The R-UCLA Loneliness Scale (3rd Version)

This scale is a twenty-item self-reported instrument, which was constructed and developed by Russell (1978; Russell, Peplau, & Cutrona, 1980) and used globally in diverse populations (Shahidi, 2013; Kwiatkowska, Rogoza, & Kwiatkowska, 2017; Suri et al., 2020). This scale was introduced not only for clinical practice as a sensitive and specific instrument in detecting and monitoring the sense of loneliness during the course of treatment, but also it has been used widely in different psycho-social research areas (Bhagchandani, 2017; Hunley, 2010; Shahidi, 2013; Stickley & Koyanagi, 2016; Yildiz, 2016). The scale was created using friendly language that can be easily and quickly completed. Having different versions with either 11 items, 6 items, or 20 items, the 20-item version of R-UCLA Loneliness Scale was used in this study.

Each item implies a possible sense of loneliness in different social and interpersonal contexts caused by lack of companionship, being isolated from others, being withdrawn, or lack of empathy with others. Some examples of items include, "*I am unhappy being so withdrawn*", "*I feel left out*", "*There is no one I can turn to*", and "*I lack companionship*". Items on this scale require Likert-style responses ranging from "1= never" to "4= most of the time". A respondent's score is obtained by summing his/her scores on all 20 items.

The scale has reasonable validity and reliability (Kwiatkowska, et al., 2017; Russell et al., 1980; Suri et al., 2020; Terrell, & Dringus, 2000; Weeks, & Asher, 2012). In one study on loneliness and psychological well-being, Shahidi (2013) examined the internal consistency of the R-UCLA Loneliness Scale (20-items) responded by university students in Canada. The results revealed that the scale has an acceptable internal consistency with Cronbach's Alpha of 0.87. Dogan et al. (2011) also studied the psychometric properties of this scale. Their results demonstrated that the scale has both high internal consistency (0.96) and test-retest reliability (0.94). Likewise, Yildiz's (2016) examination of the internal consistency of the R- UCLA-LS resulted in high Cronbach alpha (0.92).

2.2.2. Differential Loneliness Scale (DLS) – Student Version

Originally designing and developing by Schmidt and Sermat in 1983 to assess loneliness in students (young

adults), this scale was psychometrically developed and used in many recent studies (Ahmed, 2019; Goossens, Maes, Danneel, Vanhalst, & Nelemans, 2017, Shahidi, 2013; Shahidi, French, Shojaee, & Bellido Zanin, 2019). The scale measures loneliness in four types of relationships, namely romantic/marital (sexual) relationships (R/M), friendships (Fr), relationships with family (Fam), and relationships with larger groups (Gr). DLS has good validity and reliability making it appropriate for research (Schmidt, & Sermat, 1983). The person completing the measure reads each 20 item stems and then indicates whether it is True or False. Blazin, Settle, and Eddins (2008) demonstrated that the scale has acceptable reliability for research. In one study conducted on the same population, Shahidi et al. (2019) demonstrated that DLS had a reasonable internal consistency with Cronbach's Alpha 0.70.

2.2.3. Ryff's Psychological Well-Being Scale (PWBS)

This scale was used to examine the discriminant validity of R-UCLA Loneliness Scale. The PWB scale measures six components of psychological well-being including self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life and personal growth (Cheng, & Chan, 2005; McDowell, 2010; Ryff, 1989, 2013, 2017; Van Dierendonck, 2005). From different versions, the 42-item scale was used in this study. The scale had reasonable validity and reliability examined by several researchers in different societies (Van Dierendonck, 2005; Ryff, 1989; Ryff, & Singer, 2008; McDowell, 2010; Shahidi, 2013). In the target population (IAU-TCB), two primary psychometric studies were done on 42-item version of Ryff's PWB scale demonstrating Cronbach's Alpha 0.88 (Fattahi, 2016) and 0.87 (Rezaghan, 2018).

2.3. Procedure

The questionnaires were used in this study were original based on receiving permissions from their owners (authors). The scales were translated into Persian and back translated to English by two experts both were fluently bilingual in English and Persian. After the confirmation of translation accuracy and preparing the materials, the researchers pursued the ethics' principles of the university. Accordingly, all potential participants were informed, both verbally and in a written covering letter, of their rights. Participation was voluntary. Those who chose to participate could decline to answer any questions or to withdraw from the study at any time freely. After explaining the confidentiality of the research, the recruited samples were asked to complete all questionnaires.

Collected data were analyzed by "R" software to detect Differential Item Functioning (DIF) via Ordinal Logistic Regression (OLR). DIF is one of the measurement invariance methods that is used to determine whether a scale (along with its items) in different examinee groups is fair or not (Sadeghi, & Abolfazli Khonbi, 2017). The violation of measurement invariance could result in individuals receiving different scale scores systematically while experiencing the same level of loneliness (Wang, Strobl, Zaileis, & Merkle, 2018). As Holland and Thayer (1988 cited in Bulut & Suh, 2017) defined: "DIF refers to a conditional dependency between group membership of examinees (e.g., male vs. female) and item performance (i.e., the probability of answering the item correctly) after controlling for latent traits. As a result of DIF, a biased item provides either a constant advantage for a particular group (i.e., uniform DIF) or an advantage varying in magnitude and/or in direction across the latent trait continuum (i.e., non-uniform DIF; p 1-2)".

Since different groups exist in each cohort of examinees, ignoring the stability of measurement across groups can make the interpretation of results ambiguous (Wang, et al., 2018). This can increase the DIF of items (Gierl, Gotzmann, Boughton, 2004, Bulut, 2015). Thus, DIF could be a serious threat for instrument validity (Mousavi, Shojaee, Shahidi, Cui & Kutcher, 2019; Mousavi and Krishnan, 2016). Detection of DIF could be done with several methods such as Logistic Regression (LR), which was employed in this study to flag potential problematic items. Since LR does not require a large sample size and also there is no need for the specific forms of item response function, this statistical procedure is considered as a popular DIF detection method (Narayanan & Swaminathan, 1994 cited in Chen & Jin, 2018). As LR can be extended to the polytomous item response data (Sharafi, Mousavi, Ayatollahi & Jafari, 2017; Yesiltas & Paek, 2020), we limited our attention to the observed score-based DIF method utilizing Ordinal Logistic Regression.

Ordinal LR was done using RStudio, version 1.2.5001, and lordif package (Choi, Gibbons & Crane, 2011). Detecting DIF with utilizing OLR is based on comparing three different nested models. Zumbo (1999) have modeled the following forms:

$$\eta_i = ln\left[\frac{p(Yi \leq klg, \theta)}{1 - p(Yi \leq klg, \theta)}\right] = \beta_0 + \beta_1\theta + \beta_2 g + \beta_3(\theta * g)$$
$$\text{Model 3}$$

$$\eta_i = ln\left[\frac{p(Yi \leq klg, \theta)}{1 - p(Yi \leq klg, \theta)}\right] = \beta_0 + \beta_1\theta + \beta_2 g \quad \text{Model 2}$$

$$\eta_i = ln\left[\frac{p(Yi \leq klg, \theta)}{1 - p(Yi \leq klg, \theta)}\right] = \beta_0 + \beta_1\theta \quad \text{Model 1}$$

Where $p(Yi \leq k)$ is the probability of responding at or below category k to an item for the ith person, θ (person's latent trait levels, ability parameter; Hambleton, Swaminathan, & Rogers, 1991 cited in Kim & Yoon 2011) represents the overall level of loneliness and it is measured by the total test score, g is a grouping variable, and $g*\theta$ represents the interaction between the grouping variable and the overall level (Mousavi, et al. 2019). The difference value of $-2\times$ log-likelihood between model 1 and 3 can be used to detect two forms of DIF, uniform and non-uniform. This difference value is compared to a Chi-squared distribution with two degrees of freedom and significant result will flagged the item for DIF. To determine type of DIF, we

examined "item true score function plots", which illustrate the item characteristic curve (ICCs; also called item response function for polytomous items (ICF) and represents expected scores conditional on the trait level), to see if the DIF is uniform or non-uniform.

3. Results

3.1. Discriminant and Convergent Validity

To examine criterion validity of the R-UCLA Scale, the researchers employed convergent and discriminant validity. Using DLS Scale to examine convergent validity, the results revealed that all four subscales and the DLS total scores have positively significant correlation, with R-UCLA scale (between $r = 0.157$ and $r = 0.521$ -Table 1). Thus, we may conclude that these two scales measure the same construct. Similar results were cited by Shahidi (2013, 2019).

Also, the researchers used Ryff's Psychological Well-Being Scale (PWBS) to investigate discriminant validity. Displaying in Table 2, the results of correlation coefficients of Ryff's PWB scale indicated that all subscales and the total scores of Ryff's PWB scale have negatively significant correlation with R-UCLA scale (between $r = -0.221$ and $r = -0.621$).

3.2. Reliability Analysis

To obtain the internal consistency of items of R-UCLA Scale, Cronbach's Alpha coefficient was calculated resulting in a value of 0.83. The Cronbach's Alpha coefficient is an adequate and acceptable based on psychometric assumptions (Beshlideh, 2012). In addition to the Cronbach's Alpha coefficient, the split-half method was used to check the internal consistency of the scale, which was corrected using the Spearmen-Brown coefficient. The results were 0.73 for part one, 0.70 for the second part and, also, correlation of 0.80 between two forms demonstrating reasonable psychometric properties of the scale.

3.3. Differential Item Functioning (DIF)

We analyzed the dataset (N = 400) on a 20 items of R-UCLA Loneliness scale for DIF related to gender. The reference and focal groups were defined as females (n = 200) and males (n = 200). All items had the same response-style with four response categories (Never, Rarely, Sometimes, & Most of the time). The instrument is constructed in a way that higher scores mean higher level of loneliness.

The likelihood ratio was employed as the DIF detection criterion at the α level of 0.01 and McFadden's pseudo R^2 was used as the effect size measure for flagged items (reference lordif package). With three iterations, "lordif" package terminated flagging and three items were found displaying gender related DIF: item 3 (*How often do you feel that there is no one you can turn to?*), item 15 (*How often do you feel you can find companionship when you want it?*), and item 16 (*How often do you feel that there are people who really understand you?*).

Graphs in Figure 1 indicate the difference in score between using scores that account DIF and those ignore for DIF. In the box plot on the left, interquartile range, representing the middle 50% of the differences (bound between the bottom and top of the shaded box), range roughly from -0.03 to +0.02 with a median of approximately -0.01. In the right graph the same difference scores are represented against the initial scores ignoring DIF ("initial theta"; Choi, et al. 2011), for both genders. Guidelines are plotted at 0.0 (solid line), that is to say no difference, and the mean of the differences (dotted line), however, in this plot these two lines are overlapped. The positive values to the left of this graph revealed that in some cases, accounting for DIF led to slightly higher scores (i.e., naive score ignoring DIF minus score accounting for DIF > 0, so accounting for DIF score is less than the naive score) for those with lower levels of loneliness, notably this appears to happen more for female individuals. The negative values to the right of this graph indicate that for those with higher levels of loneliness, accounting for DIF led to slightly lower scores, special for female individuals.

Table 1. Correlations between R-UCLA Scale and DLS Scale and its Subscales

	DLS Family Relationship	DLS Romantic/Sexual Relationship	DLS Relationship with Large Group	DLS Friendship	DLS Total
Total R-UCLA	.34**	.41**	.16**	.42**	.52**

Note. N = 400; **p < 0.01

Table 2. Correlations between R-UCLA Scale and Ryff Well-being Scale and its Subscales

	Ryff Autonomy	Ryff Positive Relationship	Ryff Environmental Mastery	Ryff Personal Growth	Ryff Purpose in Life	Ryff Self-Acceptance	Ryff Total
Total R-UCLA	-.22**	-.62**	-.36**	-.30**	-.37**	-.38**	-.56**

Note. N = 400; **p < 0.01

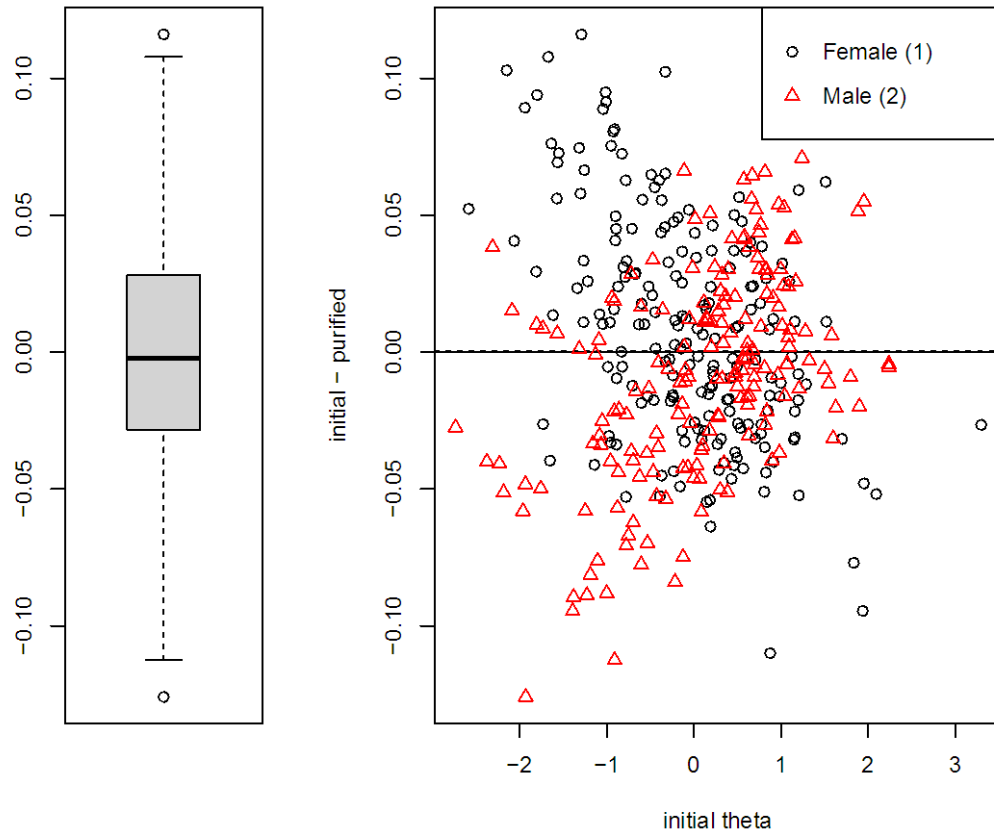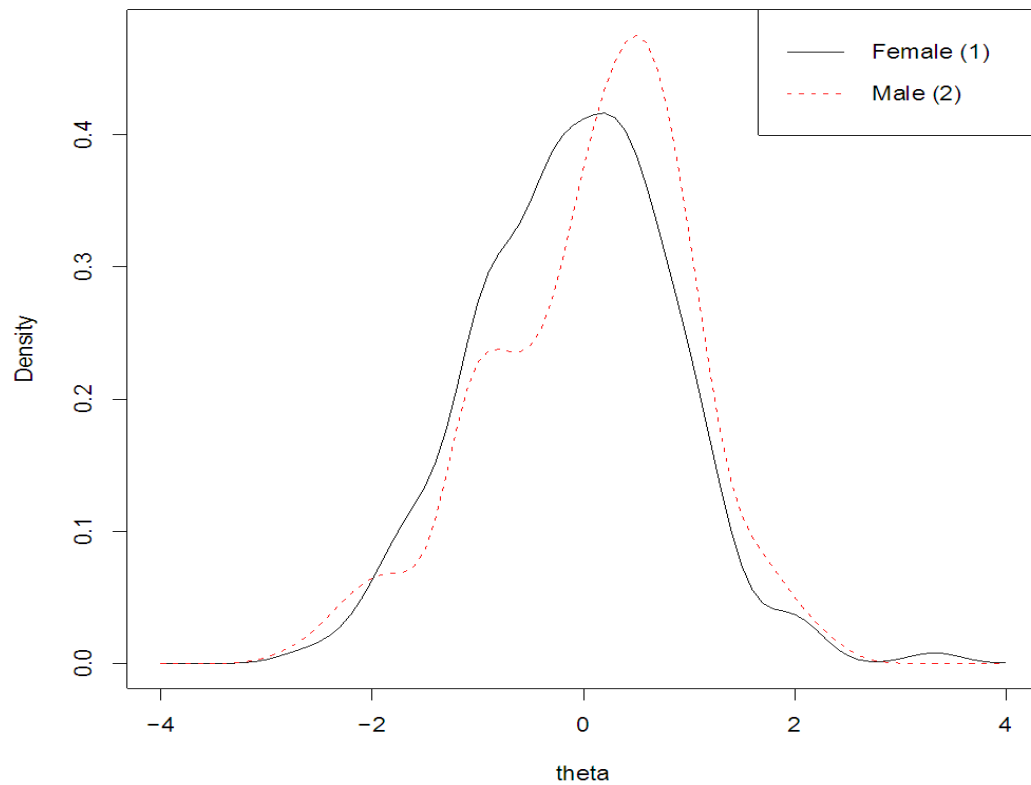**Figure 1.** Individual-Level DIF Impact

Figure 2. Trait Distributions – Females vs Males

As Figure 1 shows, the initial naive theta estimates, and the "purified" theta estimates were compared with each other through the final run of accounting DIF. The median difference over all examinees was shown in the Box-&Whisker plot, which is about 0.00 and the differences ranged from -0.135 to +0.125 with a mean of -0.01. The right graph shows, females who had lower levels of loneliness scored slightly higher than those that had higher levels of loneliness. Unlike females, males' scores grow with their levels of loneliness.

The trait distribution plot in Figure 2 shows that the latent trait in females has a more normal and bell shape than males who had more fluctuation across different theta values. This graph indicates smoothed histograms of the loneliness levels of males (dashed line) and females (solid line) sample group of participants that was measured by the R-UCLA Loneliness scale (theta).

As shown in Figure 3, Item 3 (*how often do you feel that there is no one you can turn to?*) retained all four response categories (1, 2, 3, and 4) from the four-point original rating scale. The upper-left graph shows the item characteristic curves (ICCs), dashed line for males and solid line for females. The upper-right plot displays the absolute difference between the ICCs for the two groups, illustrating that the difference is increased by distancing of mean of loneliness level (theta) from both sides. The lower-left graph shows the item response functions for the two groups (females and males) based on demographic-specific item parameter estimates (slope and category threshold values by group printed on the graph, e.g., slope = 1.08, three category thresholds -1.97, 0.9, 3.8 for females). The absolute difference between ICCs (the upper-right graph) is shown in the lower-right graph, which it weighted by the score distribution for the focal group, i.e., male individuals (dashed curve in Figure 2), revealing minimal impact.

For each flagged item, four diagnostic plots are displayed (see Figures 3-4). Based on group-specific item parameter estimates, the top left plot in Figure 3 shows item true-score functions for the item 3. The slope of the female group was higher than that for male group indicating a non-uniform DIF. Comparing Model 1 and Model 2, the LR x^2 test for uniform DIF was not significant ($p = 0.567$), while the *1-df* test for comparing Model 2 and 3, ($p = 0.001$) and *2-df* test for comparing Model 2 and 3, ($p = 0.006$) were significant.
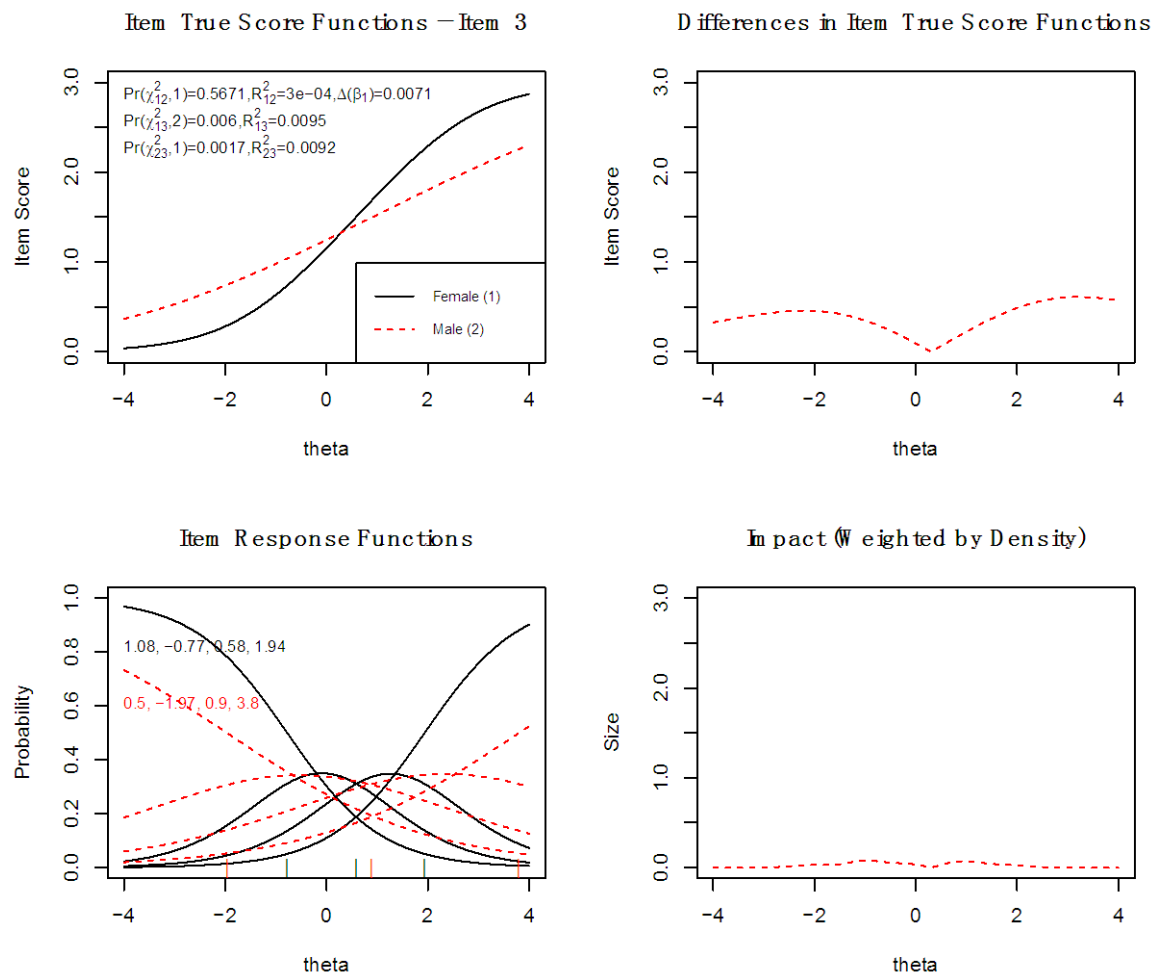

Figure 3.   Visualizing the Non-Uniform DIF for the Item # 3 in Terms of Gender
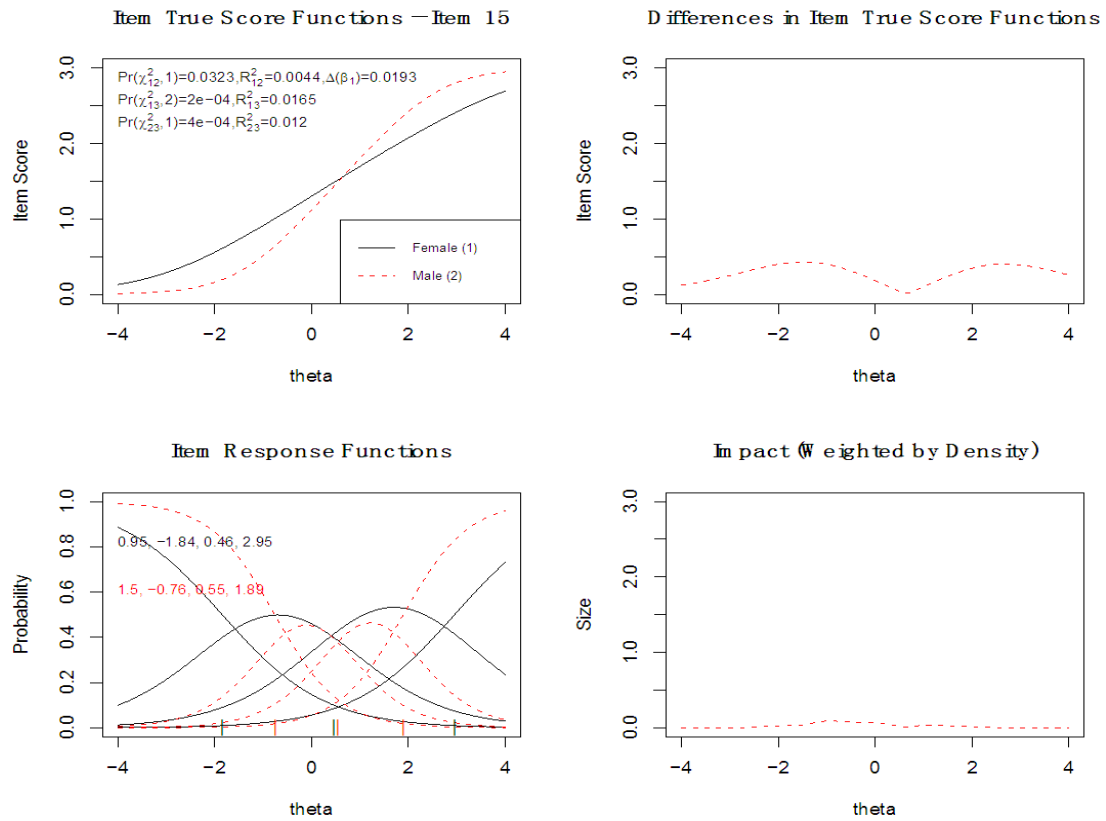
Item True Score Functions — Item 15

$Pr(\chi^2_{12}, 1)=0.0323, R^2_{12}=0.0044, \Delta(\beta_1)=0.0193$
$Pr(\chi^2_{13}, 2)=2e-04, R^2_{13}=0.0165$
$Pr(\chi^2_{23}, 1)=4e-04, R^2_{23}=0.012$

Female (1)
Male (2)

Differences in Item True Score Functions

Item Response Functions

0.95, −1.84, 0.46, 2.95
1.5, −0.76, 0.55, 1.89

Impact (Weighted by Density)

Figure 4. Visualizing the Non-Uniform DIF for the Item # 15 in Terms of Gender

Item True Score Functions — Item 16

$Pr(\chi^2_{12}, 1)=0.6681, R^2_{12}=2e-04, \Delta(\beta_1)=0.0033$
$Pr(\chi^2_{13}, 2)=0.002, R^2_{13}=0.0119$
$Pr(\chi^2_{23}, 1)=5e-04, R^2_{23}=0.0117$

Female (1)
Male (2)

Differences in Item True Score Functions

Item Response Functions

1.56, −0.96, 0.5, 2.13
2.35, −0.57, 0.34, 1.56
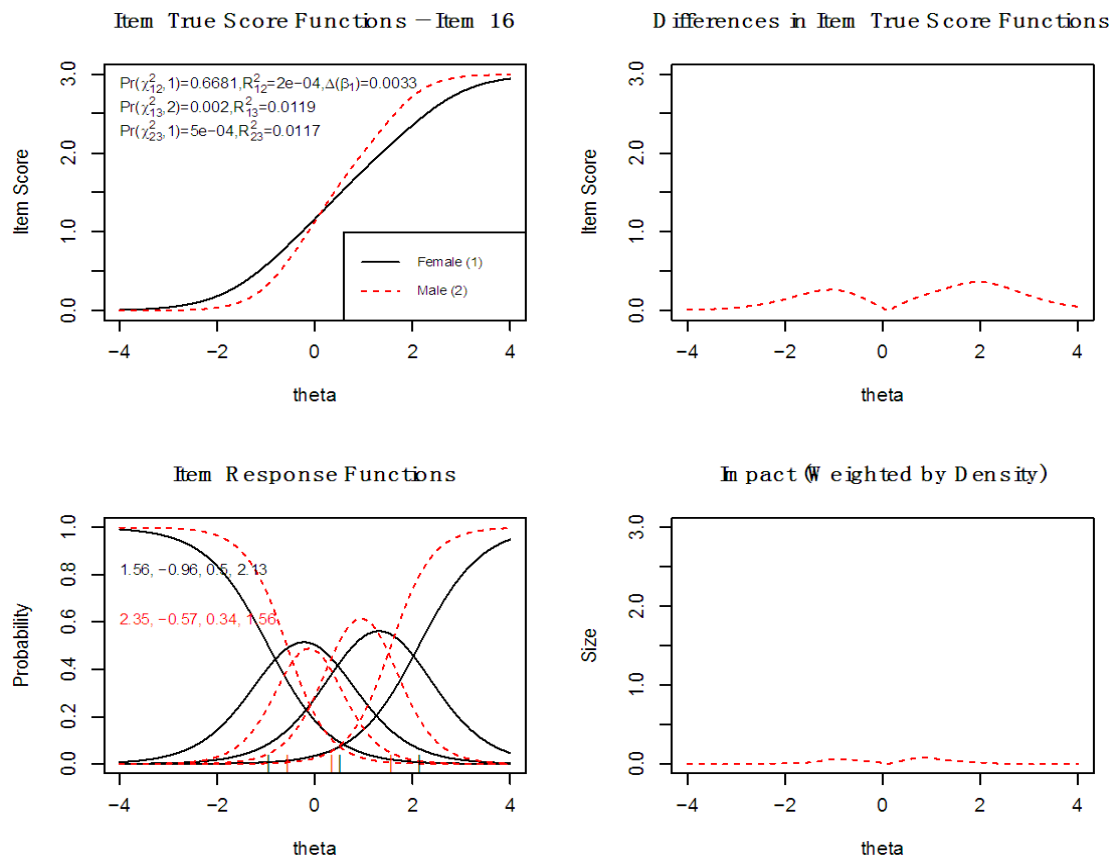
Impact (Weighted by Density)

Figure 5. Visualizing the Non-Uniform DIF for the Item # 16 in Terms of Gender

The top right plot in Figure 3 indicates the expected impact of DIF on scores with the absolute difference between the item true-score functions (Kim, et al., 2007 cited in Choi, et al., 2011). There are differences in the item true-score functions in two spots around $\theta = 2.00$ and $\theta = 2.90$, but the density-weighted impact, displayed in the bottom right plot, is negligible since few cases in this sample have that trait level. When weighted by the focal group trait distribution, the expected impact became negligible, which is also apparent in the small McFadden's pseudo R^2 as magnitude measures, printed on the top left plot. The bottom left plot in Figure 3 collocates the item response functions for male and female individuals. The difference of estimated slop parameters (1.08 vs. 0.5), printed in the graph, can also revealed the non-uniform component of DIF.

Figure 4 illustrates the plots for item # 15, *"How often do you feel you can find companionship when you want it?"*, which shows statistically significant a non-uniform DIF. Here the difference between males and females appears to be at both higher and lower levels of loneliness. As the LR χ^2_{12} was non-significant, this result suggests the DIF was primarily non-uniform. The item response functions suggest that non-uniform DIF was due to the first and last category threshold values for the focal group being smaller than that for the reference group (-1.84 vs. -0.76 & 2.95 vs. 1.89).

Figure 5 shows a slightly non-uniform DIF for the item 16 *"How often do you feel that there are people who really understand you?"*, similar to item 15, shows statistically significant a non-uniform DIF, and the LR x^2_{12} was non-significant. Here the differences between male and female individuals are across almost the entire spectrum of loneliness measured by the test. McFadden's R^2 change for non-uniform DIF was 0.011, which is considered a negligible effect size (Penfield, Gattamorta, & Childs, 2009; Cohen, 1996). Unlike the first threshold characteristic curves (printed on the left bottom graph), the item response functions show that due to be non-uniform DIF, the category threshold parameters for focal group of two last thresholds are smaller than the reference group.

In Figure 6, graphs display test characteristic curves (TCCs) for male and female individuals using demographic-specific item parameter estimates. The expected total scores for a group of items are indicated by TCCs at each loneliness level (theta). These curves have two suggestions: First, there are minimal differences in the total expected score at the overall test level, and second, the differences are at a higher loneliness level (theta) for males and females in this population study. The left plot in the Figure 6 is based on item parameter estimates for all 20 items. The right plot is only about group-specific parameter estimates, DIF items (items 3, 15, and 16). Even though the recognized impact in the right plot is not very notable, the difference in the TCCs implies that males would score lower (i.e., less loneliness) if gender-specific item parameter estimates were used for scoring. Aggregating over all the items in the test and also cancelling of differences in opposite directions, sometimes differences in item functioning may become negligibly small (Figure 6, left panel), and that appears to be what happened here.
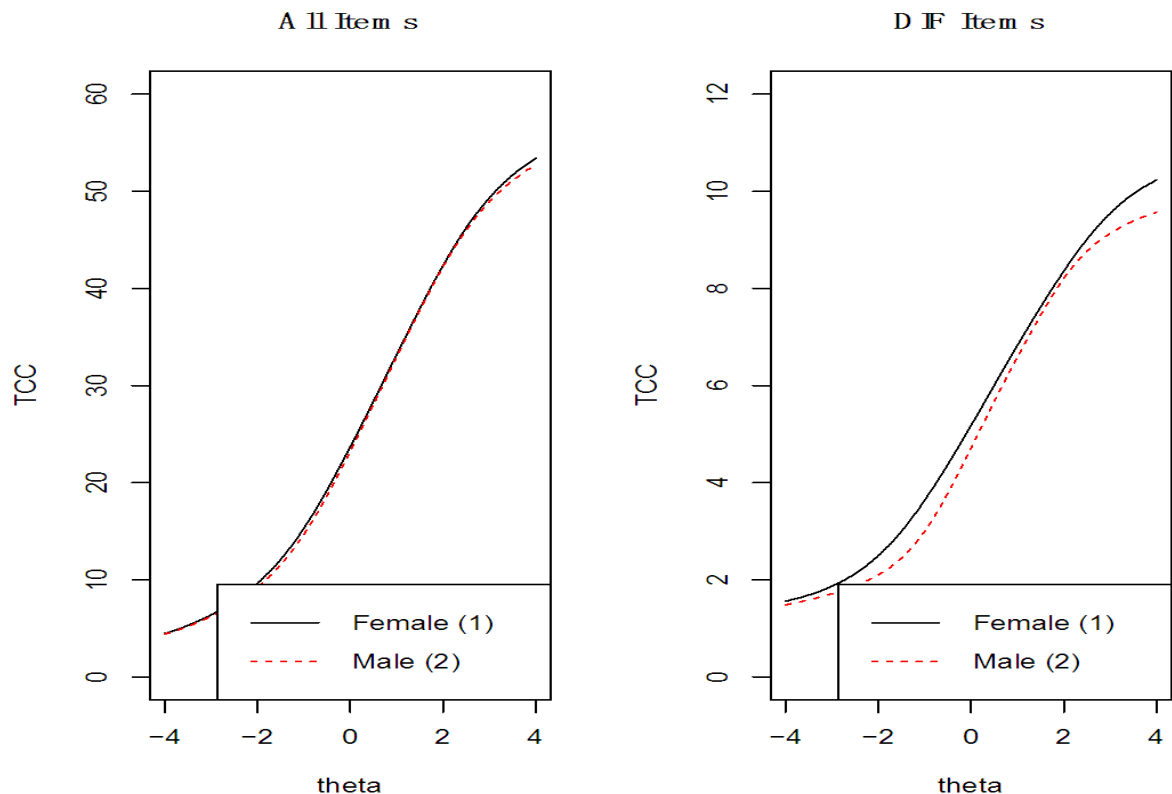

Figure 6. The Impact of DIF Items on Test Characteristic Curves

4. Discussion and Conclusions

The previous literature implies diverse relationships between loneliness and mental health problems and the necessity of using reliable and valid measurement measures (Wiseman, et al., 2006; Wakefield, et al., 2020). Evaluating loneliness as a psychological phenomenon is especially complicated since psychological constructs cannot be entirely quantified using standard and laboratory procedures (Russell & Pang, 2016). Accordingly, this not only implies the requirement of introducing valid and reliable tools, but it is also essential to have evidence-based studies regarding the fairness of measurement as stated by the AERA/APA/NCME Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The current debates on the revised standards for educational and psychological testing revealed that fairness of measurement is a gold standard for any types of psychological scales (Jiménez García, 2017; Jonson, Trantham, & Usher, 2019).

In accordance with previous studies (Shahidi et al., 2019; Yildiz, 2016) the current research revealed that R-UCLA Loneliness Scale has reasonably acceptable convergent and discriminant validity and internal consistency in diverse populations. However, to date, there is no study that have focused on the validity of R-UCLA Loneliness Scale with respect to measurement invariance and potential bias in relation to gender. Regarding the R-UCLA Loneliness Scale with Likert type items, this study used Ordinal Logistic Regression in order to assess differential item functioning of R-UCLA Loneliness Scale items and differential test functioning across gender. The psychometric properties of the R-UCLA Loneliness Scale were also examined as a prerequisite for DIF analysis. Also, unidimensionality of the scale (Shahidi, 2013; Suri, et al., 2020), showed the appropriateness to analyze DIF.

Scrutinizing the measurement invariance of R-UCLA Loneliness Scale revealed that three items (3, 15, and 16) showed gender-related DIF; that is, responses to these items were dependent on respondent's gender. Although these items are not fair based on gender, the examination of effect sizes suggested that observed DIF is practically negligible. Thus, we can conclude that the R-UCLA Loneliness Scale has validity evidence related to measurement invariance across gender therefore suitable for using in both male and female populations. However, we should be cautious when interpreting the results of DIF across gender. In the current research there were some limitations for consideration. The first one is related to the nature of samples who were not recognized having loneliness or other psychological problems. Thus, we recommend future researchers to focus on clinical population with balanced number of male and female samples. Although the sample size was enough for doing the analysis, we recommend increasing sample size to minimize the type-I error rate and maximize the chance of generalizability. Also, we recommend considering other influential factors for analyzing the fairness and validity including doing research on marital status, citizenship (rural and urban citizens), level of education and different disciplines. In conclusion, the current study was a significant step towards providing theoretical and practical information regarding the assessment of loneliness among young adults showing satisfactory evidence for R-UCLA Loneliness Scale fairness across gender.

Declaration of Conflicting Interest

'Declarations of interest: none'.

Funding

REFERENCES

[1] Ahmed, O. (2019). Psychometric assessment of the Bangla UCLA Loneliness Scale-Version 3. *Balngladesh Journal of Psychology*, 22,35-53. Retrieved from https://www.research gate.net/publication/339434492_Psychometric_Assessment_ of_the_Bangla_UCLA_Loneliness_Scale_-_version_3.

[2] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

[3] Bhagchandani, R. K. (2017). Effect of loneliness on the psychological well-being of college students. *International Journal of Social Science and Humanity, 7*(1). 60-65. https://doi.org/10.18178/ijssh.2017.7.1.796.

[4] Bjorner JB, Smith KJ, Orlando M, Stone C, Thissen D, Sun X (2006). *IRTFIT: A Macro for Item Fit and Local Dependence Tests under IRT Models.* Quality Metric Inc, Lincoln, RI.

[5] Blazin, C., Settle, A. G., & Eddins, R. (2008). Gender role conflict and separation-individuationdifficulties: their impact on college men's loneliness. *The Journal of Men's Studies, 16*(1), 69-81. https://doi.org/10.3149/jms.1601.69.

[6] Bulut, O., (2015). Applying item response theory models to entrance examination for graduate studies: practical issues and insights. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi 6*(2). https://doi.org/10.21031/epod.17523.

[7] Bulut, O., & Suh, Y. (2017). Detecting multidimensional differential item functioning with the multiple indicators multiple causes model, the item response theory likelihood ratio test, and logistic regression. *Frontiers in Education,* 2:51. https://doi.org/10.3389/feduc.2017.00051.

[8] Buchholz, E. S., & Catton, R. (1999). Adolescents' perceptions of aloneness and loneliness. *Adolescence, 34,* 203 – 213. http://www.ncbi.nlm.nih.gov/pubmed/10234378/.

[9]   Cacioppo, J. T., & Hawkley, L. C. (2009). Loneliness. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior*, pp. 227-240. New York, NY, US: The Guilford Press. https://psycnet.apa.org/record/2009-12071-015.

[10] Cauberghe, V., Wesenbeeck, V., De Jans, S., Hudders, L., & Ponnet, K. (2020). How adolescents use social media to cope with feelings of loneliness and anxiety during COVID-19 lockdown. *Cyberpsychology, Behavior, And Social Networking, 0(*0), https://doi.org/10.1089/cyber.2020.0478.

[11] Chen, H-F., & Jin, K-Y. (2018). Applying logistic regression to detect differential item functioning in multidimensional data. *Frontiers in Psychology, 9*: 1302. https://doi.org/10.3389/fpsyg.2018.01302

[12] Cheng, S., & Chan, A. C. M. (2005). Measuring psychological well-being in the Chinese. *Personality & Individual Differences*, *38* (6), 1307, http://dx.doi.org/10.1016/j.paid.2004.08.013.

[13] Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *Journal of Statistical Software, 39*(8), 1-30. URL http://www.jstatsoft.org/v39/i08/.

[14] Cohen, B. (1996). *Explaining Psychological Statistics*. Brooks/Cole Publishing. USA.

[15] De Ayala, R. J., Kim, S.-H., Stapleton, L. M., & Dayton, C. M. (2002). Differential Item Functioning: A Mixture Distribution Conceptualization. *International Journal of Testing, 2*(3-4), 243-276. https://doi.org/10.1080/15305058.2002.9669495.

[16] Dogan, T., Çötoka, N. A., Tekin, E. G. (2011). Reliability and validity of the Turkish version of the UCLA Loneliness Scale (ULS-8) among university students. *Procedia Social and Behavioral Sciences*, 15; 2058–2062. https://doi.org/10.1016/j.sbspro.2011.04.053.

[17] Ercole, V., & Parr, J. (2020). Problems of loneliness and its impact on health and well- being. *Springer Nature Switzerland AG*, https://doi.org/10.1007/978-3-319-69627-0_127-1.

[18] Fattahi, M. (2016). *A study on Ryff's PWB scale in university students*. BA's thesis research. Islamic Azad University, Tehran Central Branch.

[19] Gierl, M. J., Gotzmann, A., & Boughton, K. A. (2004). Performance of SIBTEST when the percentage of DIF items is large. *Applied Measurement in Education, 17*, 241–264. https://doi.org/10.1207/s15324818ame1703_2.

[20] Goossens, L., Maes, M., Danneel, S., Vanhalst, J., & Nelemans, S.A. (2017) Differential Loneliness Scale. In: Zeigler-Hill V., Shackelford T. (eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_23-1.

[21] Groarke, J. M., Berry, E., Graham-Wisener, L., McKenna-Plumley, P. E., McGlinchey, E., Armour, C. (2020). Loneliness in the UK during the COVID-19 pandemic: Cross-sectional results from the COVID- 19 Psychological Wellbeing Study. *PLoS ONE, 15*(9): e0239698. https://doi.org/10.1371/journal.pone.0239698, https://doi.org/10.1001/jama.2020.9740.

[22] Holmes, E. A., O'Connor, R.C., Perry, V. H., Tracey, I., Wessely, S., …, Arseneault, L. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. Lancet Psychiatry. 1; 7(6): 547–60. https://doi.org/10.1016/S2215-0366(20)30168-1.

[23] Hunley, H. A. (2010). Students' functioning while studying abroad: The impact of psychological distress and loneliness. *International Journal of Intercultural Relations, 34*(4), 386-392. https://doi.org/10.1016/j.ijintrel.2009.08.005.

[24] Igbokwe, C. C., Ejeh, V. J., Agbaje, O.S., Umoke, P.I., Iweama, C.N., & Ozoemena, E. L. (2020). Prevalence of loneliness and association with depressive and anxiety symptoms among retirees in Northcentral Nigeria: a cross-sectional study. *BioMedical Central Geriatrics,* 20:153. https://doi.org/10.1186/s12877-020-01561-4.

[25] Jaremka, L. M., Fagundes, C. P., Glaser, R., Bennett, J. M., Malarkey, W. B., & Kiecolt-Glaser, J. K. (2013). Loneliness predicts pain, depression, and fatigue: Understanding the role of immune dysregulation. *Psychoneuroendocrinology, 38*(8), 1310-1317. https://doi.org/10.1016/j.psyneuen.2012.11.016.

[26] Jiménez García, E. (2017). Standards for Educational and Psychological Testing. *Revista Española de Pedagogía, 75*(266), 170–171. https://reunir.unir.net/handle/123456789/6198.

[27] Jonson, J. L., Trantham, P., & Usher, T. B. J. (2019). An evaluative framework for reviewing fairness standards and practices in educational tests. *Educational Measurement: Issues & Practice, 38*(3), 6–19. https://doiorg.ezproxy.msvu.ca/10.1111/emip.12259.

[28] Kaiser Family Foundation; The Economist. (2020). *Survey on Loneliness and Social Isolation in the United States, the United Kingdom, and Japan.* Retrieved from http://files.kff.org/attachment/Topline-Kaiser-Family-Foundation-The-Economist-Survey-on-Loneliness-and-Social-Isolation-inthe-United-States-the-United-Kingdom-and-Japan.

[29] Keefe, J., Andrew, M., Fancey, P., & Hall, M. (2006). *Final Report: A Profile of Social Isolation in Canada.* Retrieved from https://www.health.gov.bc.ca/library/publications/year/2006/keefe_social_isolation_final_report_may_2006.pdf.

[30] Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*:2, Pp. 212-228. https://doi.org/10.1080/10705511.2011.557337.

[31] Kwiatkowska, M. M., Rogoza, R., & Kwiatkowska, K. (2017). Analysis of the psychometric properties of the Revised UCLA Loneliness Scale in the Polish adolescent sample. *Current Issues in Personality Psychology*, *5*, 1–7. https://doi.org/10.5114/cipp.2017.69681.

[32] Layden, E. A., Cacioppo, J. T., & Cacioppo, S. (2018). Loneliness predicts a preference for larger interpersonal distance within intimate space. *PLOS ONE, 13*(9), Pp. 1-21. https://doi.org/10.1371/journal.pone.0203491.

[33] Masi, C. M., Chen, Y. H., Hawkley, L. C & Cacioppo, J.T. (2011). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review 15*(3), 219–66. sagepub.com/journalsPermissions.nav https://doi.org/10.1177/1088868310377394 http://pspr.sagepub.com.

[34] McDowell, I. (2010 Back). Measures of self-perceived

well-being. *Journal of Psychosomatic Research, 69* (1), 69-79. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20630265.

[35] McGinty, E. E., Presskreischer, R., Han, H., & Barry, G. L. (2020). Psychological distress and loneliness reported by US adults in 2018 and April 2020. *American Medical Association, 324*(1), pp. 93-94. https://jamanetwork.com/journals/jama/fullarticle/2766941.

[36] McWhirter, B. T. (1990). Loneliness: A review of current literature, with implications for counseling and research. *Journal of Counseling & Development, 68*; 417-423. https://doi.org/10.1002/j.1556-6676.1990.tb02521.x.

[37] Mousavi, A., Shojaee, M., Shahidi, M., Cui, Y., & Kutcher, S. (2019). Measurement invariance and psychometric analysis of Kutcher Adolescent Depression Scale across gender and marital status. *Journal of Affective Disorders,* 253, pp. 394–401. https://doi.org/10.1016/j.jad.2019.05.010.

[38] Mousavi, A., & Krishnan, V., (2016). Measurement invariance of early development instrument (EDI) domain scores across gender and ESL status. Alta. J. Educ. Res. 62 (3), 288–305. https://cjc-rcc.ucalgary.ca/index.php/ajer/article/view/56237.

[39] Mullins, L. C. (2007). Loneliness. In J. E. Birren (Ed.), Encyclopedia of Gerontology (Second Edition), (pp. 93-98). Elsevier. ISBN 9780123708700, https://doi.org/10.1016/B0-12-370870-2/00115-3.

[40] Penfield, R. D., Gattamorta, K., & Child, R. A. (2009). An NCME instructional module on using differential step functioning to refine the analysis of DIF in polytomous items. *Educational Measurement: Issues and Practice, 28*(1), https://doi.org/10.1111/j.1745-3992.2009.01135.x.

[41] Rezaghan, S. (2018). Primary validation of Ryff's PWB scale. BA's thesis research. Department of Psychology, Islamic Azad University, Tehran Central Branch.

[42] Revicki, D. A., Chen, W.-H., & Tucker, C. (2015). *Developing item banks for patient-reported health outcomes.* In S. P. Reise & D. A. Revicki (Eds.), *Multivariate applications series. Handbook of item response theory modeling: Applications to typical performance assessment* (p. 334–363). Routledge/Taylor & Francis Group.

[43] Russell, D. (1996). The UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment, 66*, 20–40. https://doi.org/10.1207/s15327752jpa6601_2.

[44] Russell, D. W., McRae, C., & Gomez, M. (2012). Is loneliness the same as being alone? *Journal of Psychology, 146*(1–2), 7–22. https://doi.org/10.1080/00223980.2011.589414.

[45] Russell, D. W., Peplau, L. A., & Ferguson, M. L. (1978). Developing a measure of loneliness. *Journal of Personality Assessment, 42*, 290–294. https://doi.org/10.1207/s15327752jpa4203_11.

[46] Russell D.W., Pang Y.C. (2016) Loneliness. In: Zeigler-Hill V., Shackelford T. (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham.https://doi.org/10.1007/978-3-319-28099-8_1085-1.

[47] Russell, D. W., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology, 39*(3). 472-480. http://dx.doi.org/10.1037/0022-3514.39.3.472.

[48] Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069. Retrieved from https://pdfs.semanticscholar.org/7eb5/1dfece4f39df7c5c3aefa1276ae1116473a5.pdf.

[49] Ryff, C.D. (2013). *Eudaimonic well-being and health: Mapping consequences of self-realization* In A. S. Waterman (Ed.). The best within us: Positive psychology perspectives on eudaimonia, 77-98. Washington, DC US: American Psychological Association.

[50] Ryff, C. D. (2017). Eudaimonic well-being, inequality, and health: Recent findings and futuredirections. *International Review of Economy, 64,* 159–178. Retrieved from http://aging.wisc.edu/pdfs/4326.pdf.

[51] Ryff, C. D., & Singer, B. H. (2008). Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of Happiness Studies*, *9* (1), 13. https://doi.org/10.1007/s10902-006-9019-0.

[52] Sadeghi, K., & Abolfazli Khonbi, Z., (2017). An overview of differential item functioning in multistage computer adaptive testing using three-parameter logistic item response theory. *Language Testing in Asia,* 7: 7. https://doi.org/10.1186/s40468-017-0038-z.

[53] Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores." *Psychometrika Monograph, 17.*

[54] Schmidt, N. & Sermat, V. (1983). Measuring loneliness in different relationships. *Journal of Personality and Social Psychology*, *44*, 1038-1047.

[55] Shahidi, M. (2013). Loneliness as a predictor of mental health components. Master's Thesis in Child and Youth Studies. Mount Saint Vincent University, Halifax, NS, Canada. http://ec.msvu.ca:8080/xmlui/handle/10587/1275.

[56] Shahidi, M., French, F., Shojaee, M., & Bellido-Zanin, G. (2019). Predicting students' psychological well-being through different types of loneliness. *International Journal of Clinical Psychiatry, 7(1), 8-17.* doi: 10.5923/j.ijcp.20190701.02, http://article.sapub.org/10.5923.j.ijcp.20190701.02.html.

[57] Sharafi, Z., Mousavi, A., Ayatollahi, S. M. T., & Jafari, P. (2017). Assessment of differential item functioning in health-related outcomes: A simulation and empirical analysis with hierarchical polytomous data. *Computational and Mathematical Methods in Medicine,* 11. https://doi.org/10.1155/2017/7571901.

[58] Smith, K. J. & Victor, C. (2019). Typologies of loneliness, living alone and social isolation, and their associations with physical and mental health. *Ageing and Society, 39;* 1709-1730. https://doi.org/10.1017/S0144686X18000132.

[59] Stickley, A., & Koyanagi, A. (2016). Loneliness, common mental disorders and suicidal behavior: Findings from a general population survey. *Journal of affective disorders, 197*, 81-87. https://doi.org/10.1016/j.jad.2016.02.054.

[60] Suri, S., & Garg, S., (2020). Psychometric properties of the UCLA loneliness scale (version 3) in Indian context. *Shodh Sarita, 7*(25). Pp. 164-169. https://www.researchgate.net/pro

file/sushma_suri/publication/346347325_psychometric_prop erties_of_the_ucla_loneliness_scale_version_3_in_indian_c ontext/links/5fbea203458515b7976f1291/psychometric-prop erties-of-the-ucla-loneliness-scale-version-3-in-indian-conte xt.pdf.

[61] Terrell, S. R., & Dringus, L. (2000). An investigation of the effect of learning style on student success in an online learning environment. *Journal of Educational Technology Systems, 28*(3), 231-238. https://doi.org/10.2190/R53M-BV BD-UGV5-77EH.

[62] Van Dierendonck, D. (2005). The construct validity of Ryff's scales of psychological well-being and its extension with spiritual well-being. *Personality & Individual Differences, 36* (3), 629. Retrieved from https://pdfs.semanticscholar.org/ d290/ea58f26017b6bccf544a5890e4097edb8b98.pdf.

[63] Wakefield, J.R., Bowe, M., Kellezi, B., Butcher, A., & Groeger, J. A. (2020). Longitudinal associations between family identification, loneliness, depression, and sleep quality. *British Journal of Health Psychology, 25*(1). https://doi.org/10.1111/bjhp.12391.

[64] Wang, T., Strobl, C., Zaileis, A., & Merkle, E. C. (2018). Score-based tests of differential item functioning via pairwise maximum likelihood estimation. *Psychometrika, 83*(1), 132–155. https://doi.org/10.1007/s11336-017-9591-8.

[65] Weeks, M. S. & Asher, R. S. (2012). Chapter 1 - Loneliness

in childhood: Toward the next generation of assessment and research. In J. B. Benson (Ed), *Advances in Child Development and Behavior,* 42, (pp. 1-39). https://doi.org/10 .1016/B978-0-12-394388-0.00001-0.

[66] Wiseman, H., Mayseless, O., & Sharabany, R. (2006). Why are they lonely? Perceived quality of early relationships with parents, attachment, personality predispositions and loneliness in first-year university students. *Personality and individual differences, 40*(2), 237-248. https://doi.org/10.101 6/j.paid.2005.05.015.

[67] Yesiltas, G., & Paek, I. (2020). A log-linear modeling approach for differential item functioning detection in polytomously scored items. *Educational and Psychological Measurement, 80*(1) 145–162. https://doi.org/10.1177/00131 64419853000.

[68] Yildiz, M. A. (2016). Serial multiple mediation of general belongingness and life satisfaction in the relationship between attachment and loneliness in adolescents. *Educational Sciences: Theory & Practice, 16*, 553-578. https://doi.org/10.12738/estp.2016.2.0380.

[69] Zumbo, B.D. (1999). *A Handbook on the Theory and Methods of Differential Item functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (ordinal) Item Scores.* Directorate of Human Resources Research and Evaluation, Department of National Defense, Ottawa, Canada.